文献与数据学报
第 6 卷　第 3 期　　2024 年 9 月

指标与主题双重演化视角下新兴主题预测与验证

——以“纳米药物”领域为例

邱　婧　奉国和　彭凯林　刘任铧
（华南师范大学经济与管理学院，广州　510006）

摘　要：［目的 / 意义］预测某领域新兴主题，对政府部门优化资源配置、企业调整战略方向、科研工作者跟踪研究趋势与热点具有重要的参考价值。［方法 / 过程］采用深度学习模型 BERTopic 进行主题抽取，从两个视角预测新兴主题：一是在指标演化视角下构建新颖性、成长性和影响力指标体系，并对影响力测度指标进行改进；二是在主题演化视角下根据主题相似度刻画主题演化关系图。从主题演化和资料审阅两维度对新兴主题预测的准确性和科学性进行双重验证。［结果 / 结论］基于多指标和主题的双重演化视角预测新兴主题的方法是科学有效的。

关键词：新兴主题预测　BERTopic　演化趋势　新兴性指标　纳米药物
分类号：G350
DOI：10.31193/SSAP.J.ISSN.2096-6695.2024.03.07

0 引　言

新兴主题预测是科技情报学界备受关注的热点。代表科技发展趋势的新兴主题不仅为国家科技战略规划提供关键参考，同时对企业提升科技竞争力、对研发人员探寻技术机会具有重要决策价值。欧盟 PromTech 项目致力于对科学文献的分析来定位新兴技术[1]。2023 年 3 月，北约科学技术组织发表的《科学技术 2023—2043 年趋势》报告整合来自多个国际机构关于颠覆性技术的测量工具和指标，以识别和预测未来 20 年全球新兴和颠覆性技术[2]。2023 年 6 月，世界经济论坛发布《Top10 Emerging Technologies of 2023》，预测未来三到五年内将对社会产生积极影响的十

［作者简介］邱婧（ORCID：0009-0003-6514-0346），女，硕士生，研究方向为数据要素、信息分析，Email: qiujing2021@163.com（通讯作者）；奉国和（ORCID：0000-0002-0774-1544），男，教授，博士，研究方向为数据要素、数据挖掘、信息计量学，Email:ghfeng@163.com；彭凯林（ORCID：0009-0007-6948-0582），女，硕士生，研究方向为数据要素、信息分析；刘任铧（ORCID：0009-0008-9154-3516），男，硕士生，研究方向为信息分析。

大新兴技术[3]。上述报告采取相关工具预测未来新兴技术发展趋势，对世界了解技术发展动态具有重要意义。

科技日新月异，如何从海量科技素材中找到有价值的“参考”或“答案”[4]，如何识别预测新兴技术，是我国新兴技术预布局的重要课题。《“十四五”规划和 2035 年远景目标纲要》提出要发展战略性新兴产业，瞄准人工智能、量子信息、生命健康、脑科学等前沿领域，实施一批具有前瞻性、战略性的国家重大科技项目[5]。因此，本研究以“纳米药物”领域为例进行实证分析，旨在探讨国内关于纳米药物领域研究主题的变化情况，创新性地从指标演化和主题演化两个视角预测新兴主题，并采用主题演化和资料审阅对新兴主题预测结果进行双重验证。

1 文献综述

当前关于新兴主题预测方法的视角主要包括两类：第一类是从指标视角，通过特征指标的变化预测新兴主题，基本思路是利用文献的内外部属性对新兴主题进行量化指标计算。如 Lu 等[6]基于 Long Short-Term Memory（简称 LSTM）神经网络模型构建时间衰变特征、频数持久性、社区规模及其发展潜力四个指标预测关键词未来的频数发展趋势。徐路路等[7]以基金资助项目文本为数据源，基于 First Story Detection（简称 FSD）模型构建资助金额、平均资助金额、资助时长、平均资助时间、布局空间和新颖度等指标来探测新兴主题。叶光辉等[8]设立新颖度、主题出现率、主题强度的新兴主题测量指标并构建新兴主题时间序列，分别采取 Back Propagation（简称 BP）神经网络和 Support Vector Regression（简称 SVR）对新兴主题未来三年发展趋势进行预测分析。第二类是从主题演化视角，通过文本内容的主题演化分析探测新兴主题，主要包括共词分析和主题分析等。Duan 等[9]在关键词共现网络基础上，通过链接预测和共邻接指数测度揭示领域知识融合的结构性机制并进一步构建面向领域的潜在知识融合预测和分析框架，刻画领域知识融合过程和趋势。吴胜男等[10]从网络结构演变特征和主题内容演变角度，对肺癌靶向药物领域不同生命周期阶段进行主题演化分析预测新兴主题。李海林等[11]对关键词进行重要性加权，构建基于权重相似度的共词矩阵，分析主题演化趋势。

综上，单一视角可能不足以展示各主题随时间变化的全貌，且在探测新兴主题及刻画演化趋势时可能存在技术内生性偏差。目前关于不同视角的综合研究远谈不上丰富，存在较大的创新思考空间[12]。为此，本研究以纳米药物科研论文为分析数据，构建新颖性、成长性和影响力新兴性测度指标体系，同时在系统分析和借鉴以往影响力测度方法基础上对影响力指标进行改进，从指标演化视角与主题演化视角预测新兴主题，并从主题演化和资料审阅两维度进行双重验证，使得本研究预测准确性更高。本研究的边际贡献是：①创新性地从两个视角（指标演化视角和主题演化视角）综合探测新兴主题；②丰富验证方式，在已有资料验证法基础上，融合主题演化验证方法。

2　研究设计

本文以“纳米药物”研究领域为例，依据研究思路设计实验。首先利用深度学习动态主题建模对纳米药物领域文本语料库进行挖掘，分别从指标演化视角和主题演化视角探测新兴主题。在指标演化视角下，利用文献计量和时间序列模型构建新颖性、成长性和影响力指标体系。在主题演化视角下，利用图谱分析法挖掘各时间窗口的主题，根据主题相似度刻画主题演化图，以此探测新兴主题结果。然后，利用主题演化图谱和资料审阅两维度双重验证新兴主题结果的准确性。具体研究思路如图 1 所示。

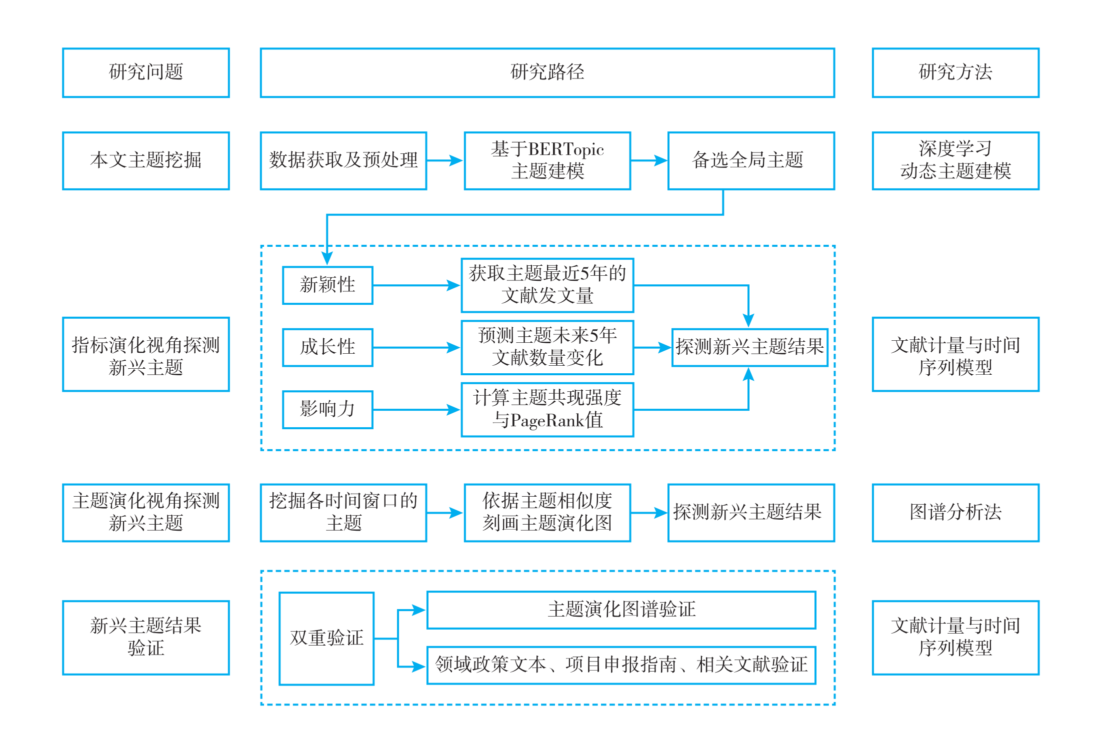


图 1　研究思路图

2.1　主题抽取

BERTopic 是一种结合 Bidirectional Encoder Representations from Transformers（简称 BERT）嵌入和基于聚类的主题建模方法。其主要步骤包括文档嵌入降低维度、主题聚类以及主题动态表示。本研究采用“all-mpnet-base-v2”预训练模型对预处理后的文本进行嵌入表示。该模型基于超过 10 亿对的训练数据，能够生成高质量的句子和段落嵌入，从而提高模型性能，减少计算资源和时间的消耗。为简化高维向量的处理，BERTopic 使用统一面域逼近和投影 Uniform Manifold Approximation and Projection（简称 UMAP）进行降维。UMAP 有效地将高维嵌入向量映射到低维空间，保留数据的全局结构特征[13]。降维后的词向量输入采用 Hierarchical Density-Based Spatial Clustering of Applications with Noise（简称 HDBSCAN）算法。该算法不仅可以减少嵌入向量，还可以对相似的文献数据进行聚类[14]，可用于自动推荐最优的簇类结果。而后利用 Class-based

Term Frequency–Inverse Document Frequency（简称 c–TF–IDF）评价每个词对每个 HDBSCAN 聚类的重要性，解释主题并在主题中保留重要词[15]。Latent Dirichlet Allocation（简称 LDA）是一种主题模型，它采用隐含的狄利克雷分布来表示文档中主题出现的概率以及每个主题中词汇出现的概率。而非负矩阵分解 Non–negative Matrix Factorization（简称 NMF）是将主题识别问题转化为约束最优化问题并通过矩阵分解的计算方法实现主题识别[16]。相较于概率主题模型 LDA 和基于矩阵分解的主题模型 NMF 方法，本研究采取 BERTopic 模型，通过自动化的超参数设置简化流程进而降低使用难度。

完成前期准备后，将纳米药物领域的研究文本语料库导入 BERTopic 模型并进行训练拟合，获取多个主题，每个主题下包含一组特征词及其对应概率。这些特征词及其概率值揭示主题的核心内容，帮助理解纳米药物领域的研究热点和发展方向。

2.2　新兴主题预测

2.2.1　从指标演化视角预测新兴主题

设计科学合理的预测指标是预测新兴主题演化趋势的基础。Rotolo 等[17]系统梳理了新兴技术相关定义，将其表述为一种具有新颖性与显著影响力，且在知识生产过程中产生的且发展相对迅速的技术。在此定义下，他们提出新兴技术主题具有激进的新颖性、增长速度更快、连贯性、影响力显著、不确定性和模糊性等特征。白敬毅等[18]使用 LDA 主题模型对科技文献进行主题划分，并构建主题新兴指标，包括主题新颖性、增长性和影响力，识别和探测新兴主题。此外，亦有许多相关研究依据这些特征构建不同指标测度主题新兴性[19-20]。因此本文构建新颖性、成长性和影响力指标进行新兴主题预测。

（1）新颖性指标。为提高新兴主题探测时效性，指定数据来源的时间跨度为 Ts 到 T（T–Ts>4），集中计算新兴特征在 T–4 到 T 期间的数值。即以 T 为临界时间点，用 T 之前五年（T–4~T）的数据预测 T 未来五年（T+1~T+5）的数据。与热门主题相比，新兴主题在早期阶段发文量相对较少，因此采用各主题在 T–4 到 T 期间的总发文量作为新颖性度量标准。鉴于衰退主题的发文量也相对较低，因此需要结合成长性指标做出评判。

（2）成长性指标。成长性指标用于评估主题在 T–4 到 T+5 期间的发文量变化趋势，以反映主题发展动向，包括增长、下降和无明显变化三类。为深入了解主题发文量演化过程，需搭建时间序列模型，以各主题 Ts 到 T 期间每年的发文量来预测 T+1 到 T+5 期间的发文量。相较于传统回归分析，这种方法在预测时专注于数值序列自身的时间变动，不受其他因素的干扰，能避免因果模型中难以满足的限定条件矛盾。

与传统时序模型相比，机器学习算法在处理异常值和强不确定性数据时预测效果更优[21]。本文采用以自动回归前馈神经网络 AutoRegressive Network（简称 AR–Net）为基础的 Neural Prophet 预测模型，面对数据缺失和趋势变化时表现出强稳健性[22]，能更有效地进行时间序列的自相关分析。采用单变量的时间序列，将时间戳和观测值输入模型进行拟合，得到未来一段时间内的预测值，再使用均方根误差 Root Mean Square Error（简称 RMSE）和平均绝对误差 Mean Absolute Error（简称 MAE）来衡量模型预测值与真实值的差距。

（3）影响力指标。主题影响力体现该主题在特定领域发展过程中的地位，展示新兴主题的

“兴”。以往研究常常采用文献被引次数高低、权利要求项数多少等衡量影响大小，但这些方法无法全面考虑主题间的复杂关系。因此，本研究改进了传统的影响力测度指标，从社会网络分析视角，选取主题共现网络特征指标，即采用 PageRank 指标测度节点在网络中的影响力。该指标在测度节点影响力时，综合考虑邻居节点数量及影响力程度[23]，以更全面地反映主题在特定领域发展过程中的地位。主题共现强度可反映不同主题间的关联程度，主要参考刘俊婉等[24]的做法：一篇文献中仅有少数主题的“文档－主题”概率较大，该文献的主题特征很可能集中于这些少数主题。因此，根据“文档－主题”概率分布，按照概率大小对每篇文献的主题进行排序，并根据二八原则，选取前 t 个累积概率约为 0.8 的主题，作为该文献的主题。主题共现强度为两个主题同时出现在多篇文献中的概率，计算公式为：

$$I_{ij}=\sum_{a=1}^{n}P\left(T_{ai}T_{aj}\right)=\sum_{a=1}^{n}P\left(T_{ai}\right)\times P\left(T_{aj}\right) \tag{1}$$

其中，I_{ij} 表示主题 i 和主题 j 的共现强度，T_{ai} 表示第 a 篇文献的主题 i，T_{aj} 表示第 a 篇文献的主题 j，n 表示主题 i 与主题 j 共现的文献数量，$P(T_{ai})$ 表示第 a 个共现文献主题 i 的概率，$P(T_{aj})$ 表示第 a 个共现文献主题 j 的概率。主题间相互独立，因此可通过条件概率推导两个主题间的共现强度。

最后，将得到的主题共现强度导入 Gephi 软件并计算各主题的 PageRank 值，以此来衡量影响力的大小。

2.2.2　从主题演化视角预测新兴主题

主题演化分析是通过文献间的关联特征对文献主题进行分析，揭示主题随时间演变的规律。当前多数研究只分析主题外部特征，未从宏观时间维度上理解和把握新兴主题演化趋势，在分析主题前后演化关系方面仍存在不足，对新兴主题出现的时间把握不准。针对此问题，本文基于演化视角，利用相邻时间段主题间相似度及演化图谱，描绘主题随时间变化呈现的出现、分化或合并等关系，预测新兴主题。

2.3　新兴主题结果验证

本研究从两个维度验证新兴主题预测的准确性和科学性。第一，利用主题演化图谱识别最新时间窗口中的潜在新兴主题；第二，利用政策文本、项目申报指南、科研论文、报告资讯等公开信息验证所预测主题，通过研读公开信息判定所预测主题的新兴性，定性评估本研究方法的准确性和实用性。

3　实证分析

3.1　数据采集与预处理

以中国知网纳米药物文献为研究数据源，确定检索式：主题 =（‘纳米药物’+‘纳米材料药物’+‘纳米药品’+‘纳米载体药物’+‘纳米药剂’+‘纳米医药’+‘纳米制药’），并且限定文

献类型为中文学术期刊和学位论文，采集题名、关键词、摘要、发表年份等重要题录信息，时间跨度为 2013—2023 年，检索时间是 2024 年 4 月 2 日，经过去重等处理，获得学术期刊论文 1 450 篇，学位论文 3 771 篇，二者合计 5 221 篇。在预处理中，合并科研论文的标题和摘要数据，以论文关键词构建自定义词典，运用 jieba 进行分词。

3.2 训练 BERTopic 模型

采用 BERTopic 模型进行主题建模，选择“all-mpnet-base-v2”作为词嵌入模型。主题数目由参数 nr_topics 确定，BERTopic 合并相似主题直至达到设定的 nr_topics 值为止。为确保主题数量合适，本文在训练模型时不具体设定 nr_topics 值，而将其设置为“auto”，由 BERTopic 自动进行迭代生成。

3.3 “纳米药物”主题结构

对 5 221 篇纳米药物领域文献运用 BERTopic 模型进行主题抽取，得到 38 个主题。提取各主题概率最大的前 5 个特征词，并据此拟定主题名称，具体如表 1。图 2 则以条形图展示部分主题概率排序前 5 的特征词。

表 1 主题及特征词分布（部分主题）

主题	主题名称	主题特征词
topic0	光学治疗技术	荧光 | 光热 | 光敏剂 | 多柔比星 | 照射
topic1	纳米粒化疗药物	纳米粒子 | 多柔比星 | 粒径 | 纳米粒 | 能力
topic2	纳米药物领域研究进展	研究进展 | 纳米技术 | 中国 | 纳米材料 | 药物制剂
……		
topic9	放射学影像与医学诊断技术	计算机断层扫描 | 磁共振成像 | 造影剂 | 核磁共振成像 | 纳米粒
……		
topic21	仿生纳米药物靶向治疗	纳米脂质体 | 动脉粥样硬化 | 透明质酸 | 秋水仙碱 | 斑块
……		
topic32	纳米药物递送系统	Hsp90 抑制剂 | 丙烯酰胺 | 脂质纳米颗粒 | 热休克蛋白抑制剂 | 靶向 NAE 的小分子抑制剂
topic33	介孔硅纳米材料	介孔硅材料 | 介孔硅纳米材料 | 介孔二氧化硅 | 桥联 | 树枝状
topic34	多肽纳米材料	多肽纳米材料 | 线粒体靶向 | 多糖 | 聚合物 – 多肽复合物 | 多肽纳米
……		

3.4 新兴主题预测

3.4.1 从指标演化视角预测新兴主题

（1）新颖性演化分析

从 BERTopic 模型得到每个主题的年度发文量如图 3 所示，揭示各主题在时间轴上的演变趋

势，以及各主题在不同年份的兴起和衰落。计算 2019 年至 2023 年各主题的总发文量，得到平均值为 55.42 篇。随后，选取发文量小于平均值 55.42 的主题作为备选主题，得到满足新颖性要求的主题共 30 个，即 topic6、topic9~topic37。

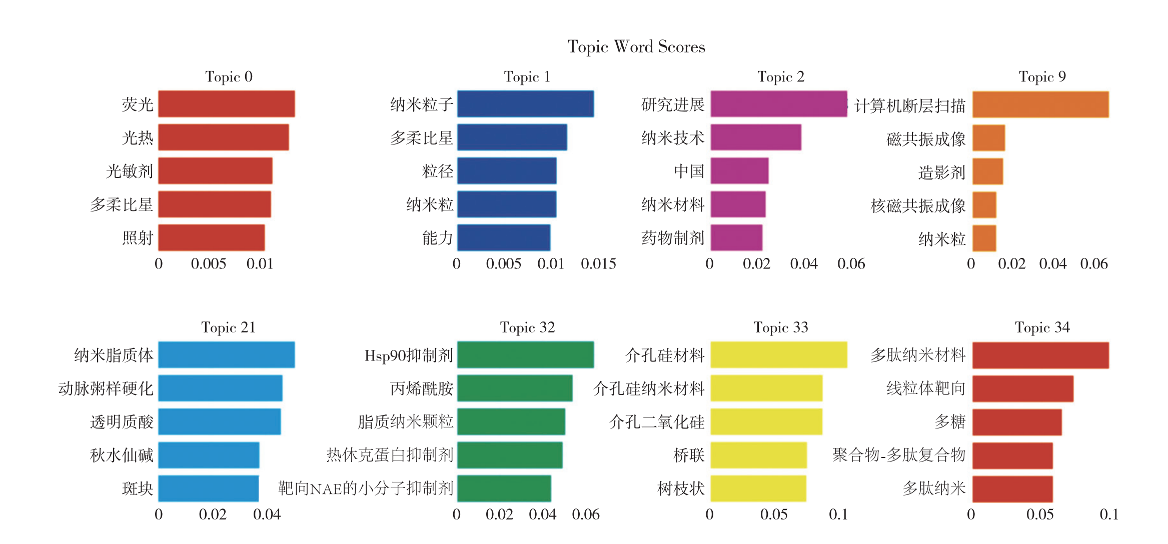


图 2　各研究主题概率排序前 5 的特征词分布（部分主题）

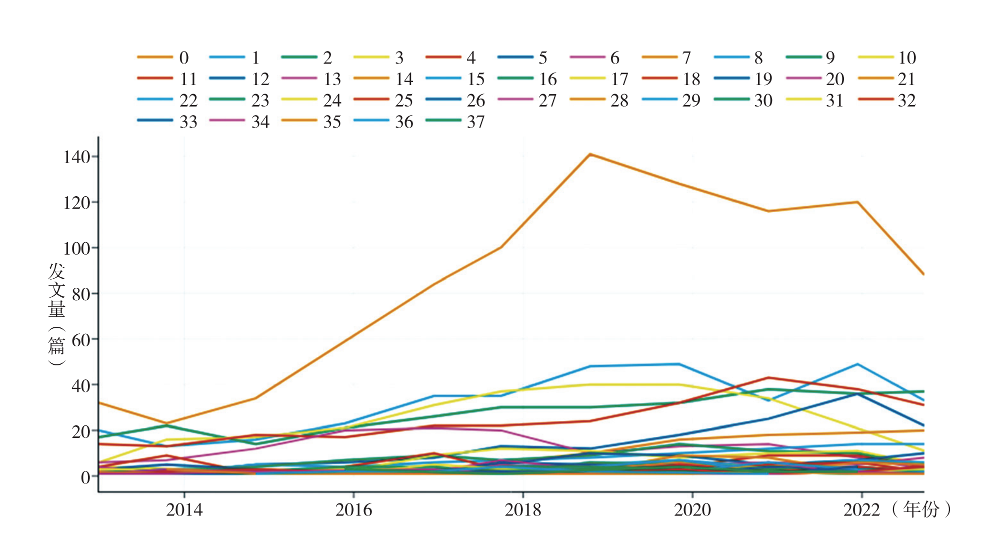


图 3　主题年度发文量变化

（2）成长性演化分析

采用 Neural Prophet 模型拟合各主题在 2013 年至 2023 年的历史发文量，预测 2024 年至 2028 年的发文量。使用均方根误差（RMSE）和平均绝对误差（MAE）两个指标来评估模型预测值与实际值之间的差异，指标值越小表示模型性能越佳。利用 Neural Prophet 模型训练和预测 38 个主题的时间序列，所得结果均较好。以主题 topic8 为例，得到的平均绝对误差和均方根误差分别是 0.144 和 0.249，两个指标值小，表明 Neural Prophet 模型拟合预测效果出色，

结果如图4所示，该图显示 Neural Prophet 模型在时间序列预测中体现主题成长性指标的演化趋势。

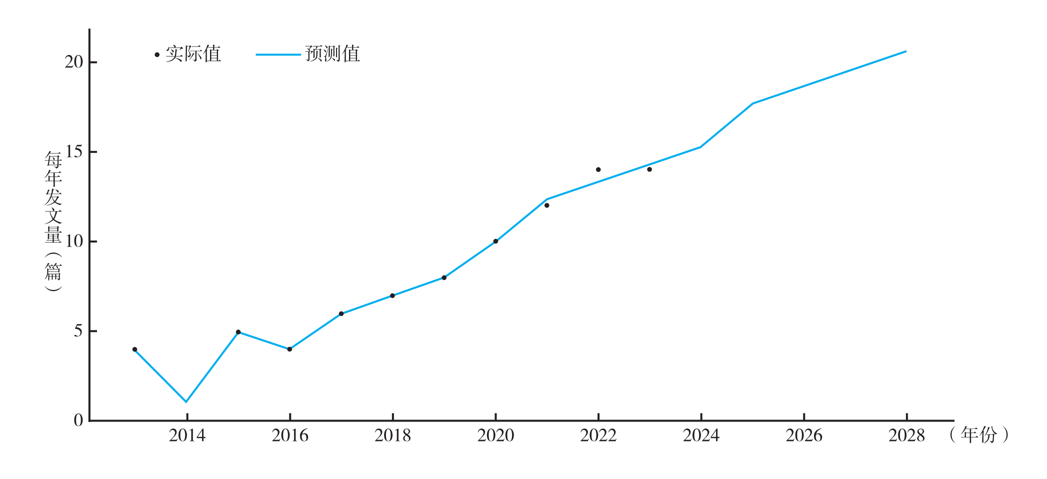


图4 Neural Prophet 模型预测曲线

按上述步骤预测其它主题在2024—2028年的发文量，发现显著上升的主题共15个，分别为topic2、topic5、topic7、topic8、topic12、topic13、topic15、topic20、topic21、topic28、topic30~topic34。

（3）影响力演化分析

根据 BERTopic 模型得到“文档—主题”概率分布，构建文档和主题之间的映射关系，绘制文档－主题概率累积分布图并标定拟合曲线，确定最能代表文档的主题数并计算主题共现强度，如图5所示。

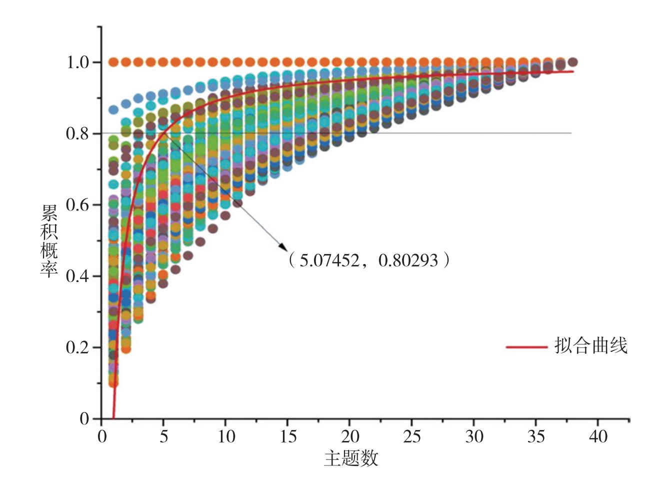


图5 文档—主题概率累积分布图

根据二八原则，取前t个累积概率约为0.8的主题作为文档主题，计算两两主题同时出现在多篇文档的概率，此概率可作为主题间的共现强度。

根据上述规则，计算得到的主题数约为5.075。选择最小的共现主题数5，即每篇文档选取

概率最大的 5 个主题形成共现对。因此，每篇文献选择其各自主题概率最大的前五个主题，这些主题较好地反映所属文献的内在主题特征。最终共获得 272 对主题共现对，表 2 展示部分主题共现强度结果，其中，weight 值为主题共现强度。

表 2　部分主题共现强度值

主题共现对	weight	主题共现对	weight
（Topic 1,Topic 4）	32.544	（Topic 5,Topic 10）	0.802
（Topic 1,Topic 10）	23.657	（Topic 5,Topic 14）	0.576
（Topic 1,Topic 19）	1.822	（Topic 5,Topic 28）	0.669
（Topic 36,Topic 10）	0.327	（Topic 0,Topic 5）	2.0834

将这 272 对主题共现对及共现强度导入 Gephi 中计算 PageRank 值，对 38 个主题的 PageRank 值加和取平均，若主题的 PageRank 值大于平均值则为有影响力的主题。符合影响力指标的主题共 8 个，分别是：topic1、topic2、topic4、topic5、topic8、topic10、topic21、topic32。

（4）基于指标的新兴主题预测

将根据新颖性、成长性和影响力分析筛选出的主题集合进行交集运算，得到 topic21、topic32 两个潜在新兴主题。根据这 2 个主题包含的特征词，结合主题最相关的文献为新兴主题命名，结果如表 3。

表 3　新兴主题预测结果

主题	主题名称	主题特征词
Topic21	仿生纳米药物靶向治疗	纳米脂质体 ׀ 动脉粥样硬化 ׀ 透明质酸 ׀ 秋水仙碱 ׀ 斑块
Topic32	纳米药物递送系统	Hsp90 抑制剂 ׀ 丙烯酰胺 ׀ 脂质纳米颗粒 ׀ 热休克蛋白抑制剂 ׀ 靶向 NAE 的小分子抑制剂

所筛选出的新兴主题具体内容如下：

① Topic21（仿生纳米药物靶向治疗）。仿生纳米药物靶向治疗是基于仿生学原理的一种先进医疗技术，它通过设计具有生物特性的纳米尺度药物载体，模拟自然生物过程以精确识别和定位疾病靶点。这种治疗方法旨在延长药物在体内的半衰期并提升稳定性和靶向能力，同时高效释放药物以达到治疗效果。其核心在于利用特定功能的纳米载体提高药物的生物利用度，减少对正常组织的影响，并降低药物剂量与给药频率，从而减轻患者的药物毒性和副作用，确保治疗的安全性和有效性。

② Topic32（纳米药物递送系统）。纳米药物递送系统是一种以纳米技术为基础，以提高肿瘤治疗效率和减少副作用的药物传输平台。它通过特定的功能性纳米载体，增强药物在体内的溶解性和稳定性，实现药物的有效载荷和保护。该系统通过延长药物在血液中的循环时间，降低生理屏障对药物的阻碍，优化药物的体内分布，并借助刺激响应机制实现智能控制释放，进而达到减少给药频率和降低副作用的目的。

3.4.2 从主题演化视角预测新兴主题

前述主题新颖性分析以 2019—2023 年为时段，为与其保持一致，此处将 2013—2023 年的数据分为三个时期：2013—2015 年、2016—2018 年、2019—2023 年。采用 BERTopic 进行主题建模，获取各个时间段的主题分布，并对相邻时间段的主题进行相似度计算。其中 2013—2015 年有 675 篇文献，2016—2018 年有 1 336 篇文献，2019—2023 年有 3 210 篇文献。

随后对各时期主题词进行相似度计算。常用相似度计算方法有夹角余弦法、皮尔逊相关系数法、杰卡德系数法[25]等，其中，夹角余弦法用于计算两个向量之间的夹角的余弦值以衡量它们的相似度；皮尔逊相关系数法用于衡量两个变量之间的线性相关性，其值介于 -1 到 1 之间；杰卡德系数法主要以乘积方式为主，通过去除分母中向量相同的部分，从而提高向量相似程度的识别度。因此，杰卡德系数法通常用于计算相近似的两个数据向量之间的相似程度。由于本研究主题词文本向量之间相近性较强，适合用杰卡德系数法进行相似度分析。根据相似度结果画出桑基图，如图 6 所示。

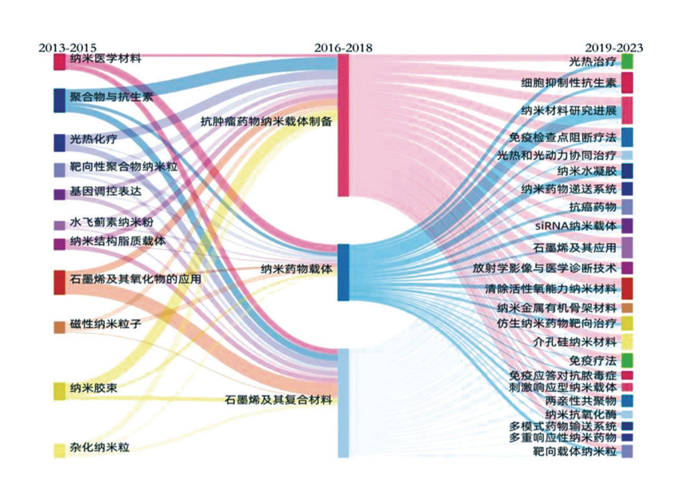


图 6 主题演化图谱

相邻时间段主题间相似度大小由线条宽度表示，宽度越大表明主题间相似度越高。从图 6 发现，“纳米药物”领域研究主题演化趋势呈现以下特点：

（1）主题数量和种类先减少后增多，各时期窗口关注的主题既多样化又相互关联。该领域主题数从 2013—2015 年的 11 个减少到 2016—2018 年的 3 个，2019—2023 年又增至 23 个。第一时期（2013—2015 年）涵盖主题有纳米医学材料、纳米粒、光热化疗、纳米脂质体、聚合物、石墨烯等方面；第二时期（2016—2018 年）涵盖主题有纳米药物载体制备、石墨烯等；第三时期（2019—2023 年）涵盖主题有纳米药物递送、纳米医学诊断技术、纳米技术治疗策略、多样化的纳米材料及研究进展、纳米粒构建等类别。从全过程来看，主题演化主要集中于药物载体、疗法、技术、材料等领域。三个时期的主题呈现丰富性与多元性，同时仍紧密围绕领域核心技术和材料展开。此外，第三时期主题内涵更趋向精细化，且此时段出现较多新兴技术和治疗方法，如纳米药物递送系统、仿生纳米药物靶向治疗、纳米金属有机骨架材料、多模式药物输送系统、

免疫检查点阻断疗法、免疫应答对抗脓毒症等。据此推断，这些技术均是“纳米药物”研究领域的研究趋势和新兴主题。

（2）石墨烯、复合物及相关应用已成为被持续关注的重点。其中，2013 至 2015 年重点在石墨烯及其衍生物的运用上；2016 至 2018 年转向石墨烯及其复合材料的技术与应用；2019 至 2023 年进一步拓宽至石墨烯及其应用领域。这一趋势反映石墨烯材料及其应用的研究在纳米药物领域具有显著影响。查阅相关纳米医学科研论文发现，现有大量研究将石墨烯材料用于药物靶向治疗[26]，作为药物载体[27]并制备金属纳米复合材料[28]等。

（3）涉及医学材料的近五年研究更精细化。例如，从第一时期（2013—2015 年）“纳米医学材料”细化到第三时期（2019—2023 年）“纳米材料研究进展、纳米水凝胶、介孔硅纳米材料、纳米金属有机骨架材料、清除活性氧能力纳米材料”等。

（4）第一时期（2013—2015 年）研究主题“聚合物与抗生素”与“光热化疗”是焦点，第二时期（2016—2018 年）关注度下降，第三时期（2019—2023 年）重新得到关注，且通过诸如“细胞抑制性抗生素”“两亲性共聚物”“光热治疗”“光热和光动力协同治疗”等具体主题，呈现出第一时期研究主题的延续与拓展。该领域进展表明，研究主题“聚合物与抗生素”与“光热疗法”在保持继承性的同时，已实现更为清晰化和精细化的发展。

3.5 结果验证

3.5.1 基于主题演化图的结果验证

根据主题演化图谱可知，2019—2023 年间纳米药物领域研究主题呈现丰富多元的特点，并向精细化发展。同时该时期还涌现出许多新兴技术和治疗方法，例如仿生纳米药物靶向治疗（topic21）和纳米药物递送系统（topic32）等。由此可知，上述技术是纳米药物领域的研究热点和趋势，与前文三个动态演化指标预测的新兴主题结果相吻合，证明本研究设计的新兴主题预测方法具有准确性和实用性。

3.5.2 基于相关资料的结果验证

利用政策文本、项目申报指南、相关研究文献等公开信息，定性评估本研究结果的准确性。

2021 年、2022 年科学技术部分别发布的“十四五”国家重点研发计划的“纳米前沿”重点专项项目申报指南中均涉及纳米科技与信息、能源、生物、医药、环境等领域的交叉研究，提出多个纳米领域的申报主题，包括通过构建仿生纳米药物体系对抗肿瘤治疗（topic21——仿生纳米药物靶向治疗）、建立药物高效递送系统（topic32——纳米药物递送系统）等。

2022 年 9 月，广州医科大学药学院附属第五医院的研究团队在期刊 *Molecular Cancer* 发表的高分被引文章指出，新型仿生纳米颗粒可靶向作用于 ADAM9、诱导癌细胞凋亡并抑制肿瘤生长和转移过程，为靶向治疗肺腺癌提供新策略[29]。2023 年 6 月，四川大学华西医院的周清华教授团队携手华西药学院的张志荣教授团队，在《药学学报》英文版发表的论文探讨了未来仿生纳米药物靶向治疗癌症临床转化的前景及挑战[30]。上述两篇论文均证明，仿生纳米药物在癌症靶向治疗方面的应用潜力（topic21——仿生纳米药物靶向治疗）和新兴价值。

Rezaie 等[31]指出基于细胞外囊泡（EVs）作为一种潜在的药物传递系统在纳米药物中得到广泛关注。Wang 等[32]指出 Fe2O3-Au 纳米药物递送系统具有 pH 响应性和磁性靶向功能，提高

药物递送的安全性和效率，为癌症治疗提供一种有前景的方法。上述研究均表明，topic32——纳米药物递送系统具有新兴性。

上述验证结果表明，本文探测的新兴主题与政策、指南、论文等资料的研究重点一致，说明本文构建的方法具有一定的合理性。

4 结 语

预测新兴主题对政府部门、企业集群、科研机构等创新主体至关重要[33]。本文以2013—2023年"纳米药物"领域的科研论文为研究对象，在指标演化视角中借助BERTopic主题模型和时间序列模型Neural Prophet模型，设计新颖性、成长性和影响力等动态演化指标以筛选潜在新兴主题；在主题演化视角中刻画主题演化图谱并预测新兴主题。通过主题演化和资料审阅两个维度对结果进行双重验证，结果表明本文提出方法有效。

本研究存在以下不足：（1）数据来源比较单一，仅考虑了科研论文数据，未考虑专利、基金等其他数据源，可能导致纳米药物领域研究主题归纳不精确；（2）对纳米药物领域的了解及主题解释有待深化。未来将在以下方面进行改进：（1）扩充实验数据，完善主题演化分析；（2）结合相关领域专家意见和权威机构判定，对所预测领域的主题进行更为全面的总结和解释。

【参考文献】

[1] Roche I, Besagni D, Franois C, et al. Identification and characterisation of technological topics in the field of molecular biology [J]. Scientometrics, 2010, 82(3): 663-676.

[2] NATO Science & Technology Organization. Science & Technology Trends 2023-2043 [R/OL]. (2023-03-23) [2024-07-31]. https://www.nato.int/nato_static_fl2014/assets/pdf/2023/3/pdf/stt23-vol1.pdf.

[3] World economic forum. Top10 emerging technologies of 2023 [R/OL]. (2023-06-26) [2024-07-31]. https://cn.weforum.org/publications/top-10-emerging-technologies-of-2023/digest/.

[4] 孙建军. 打造科技信息新生态 牢牢掌握国家科技发展主动权 [J]. 图书与情报，2021（6）：7-8.

[5] 新华社. 中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议 [EB/OL]. (2020-11-03) [2023-12-02].https://www.gov.cn/xinwen/2021-03/13/content_5592681.htm#:~:text=%E4%B8%AD%E5%8D%8E%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9B%BD%E5%9B%BD,%E5%85%B1%E5%90%8C%E7%9A%84%E8%A1%8C%E5%8A%A8%E7%BA%B2%E9%A2%86%E3%80%82.

[6] Lu W, Huang S Z, Yang J Q, et al. Detecting research topic trends by author-defined keyword frequency [J]. Information Processing&Management, 2021, 58(4):102594.

[7] 徐路路，靳杨. 基于FSD模型的政府资助项目新兴主题探测与分析 [J]. 科学学与科学技术管理，2019，40（2）：40-54.

[8] 叶光辉，王灿灿，李松烨. 基于SciTS会议文本的跨学科科研协作新兴主题识别及预测 [J]. 情报科学，2022，40（7）：126-135.

[9] Duan Y R, Guan Q. Predicting potential knowledge convergence of solar energy: bibliometric analysis based on link prediction model [J]. Scientometrics, 2021, 126(5): 3749-3773.

[10] 吴胜男，卫慧蓉，于琦，等. 结构—内容视角下的学科领域主题演化分析——以肺癌靶向药物领域为

例［J］. 信息资源管理学报，2020，10（5）：112-121.

［11］李海林，万校基，林春培. 基于关键词重要性和近邻传播聚类的主题分析研究［J］. 情报学报，2018，37（5）：533-542.

［12］岑咏华，王曰芬. 学科主题探测与演化分析：多视角述评与展望［J］. 情报学报，2023，42（4）：477-494.

［13］邱嘉萱，刘佳静，郑建明. 2023年中国图书情报领域研究概览［J/OL］. 图书馆论坛.［2024-05-28］. http://kns.cnki.net/kcms/detail/44.1306.G2.20240516.1456.004.html.

［14］Mcinnes L, Healy J, Astels S. HDBSCAN: hierarchical density based clustering［J］. The Journal of Open Source Software, 2017, 2(11): 205.

［15］Grootendorst M. BERTopic: neural topic modeling with a class based TF-IDF procedure［EB/OL］.［2024-05-28］. https://doi.org/10.48550/arXiv.2203.05794.

［16］吕璐成，周健，王学昭，等. 基于双层主题模型的技术演化分析框架及其应用［J］.数据分析与知识发现，2022，6（Z1）：18-32.

［17］Rotolo D, Hicks D, Martin B R. What is an emerging technology?［J］. Research Policy,2015,44(10): 44.

［18］白敬毅，颜端武，陈琼. 基于主题模型和曲线拟合的新兴主题趋势预测研究［J］. 情报理论与实践，2020，43（7）：130-136，193.

［19］郝雯柯，杨建林. 基于语义表示和动态主题模型的社科领域新兴主题预测研究［J］. 情报理论与实践，2023，46（2）：184-193.

［20］陈稳，陈伟. 基于计量指标多变量LSTM模型的新兴主题热度预测研究［J］. 数据分析与知识发现，2022，6（10）：35-45.

［21］马静雯，李树青，夏梦瑶. 机器学习在GDP预测中的应用研究述评［J］.科技情报研究，2022，4（3）：73-94.

［22］刘铭基，田雅楠，张亮，等.基于Prophet-ARIMA模型的民航周转量预测研究［J］. 计算机技术与发展，2022，32（2）：148-153，160.

［23］黄璐，朱一鹤，张嶷. 基于加权网络链路预测的新兴技术主题识别研究［J］. 情报学报，2019，38（4）：335-341.

［24］刘俊婉，龙志昕，王菲菲，基于LDA主题模型与链路预测的新兴主题关联机会发现研究［J］. 数据分析与知识发现，2019，3（1）：104-117.

［25］张晓琳，付英姿，褚培肖. 杰卡德相似系数在推荐系统中的应用［J］. 计算机技术与发展，2015，25（4）：158-161，165.

［26］Zhang Q, Deng S, Liu J, et al. Cancer-targeting graphene quantum dots: fluorescence quantum yields, stability, and cell selectivity［J］. Advanced Functional Materials, 2019, 29(5): 1805860.1-1805860.11.

［27］岳华，马光辉. 基于石墨烯独特生物界面效应的功能化载体研究进展［J］. 化学学报，2021，79（10）：1244-1256.

［28］徐鹏，邱汉迅，宋凌志，等. 石墨烯/金属纳米复合材料制备及研究进展［J］. 有色金属材料与工程，2017，38（3）：177-184.

［29］Liang L, Cen H , Huang J , et al. The reversion of DNA methylation-induced miRNA silence via biomimetic nanoparticles-mediated gene delivery for efficient lung adenocarcinoma therapy［J］. Molecular Cancer, 2022, 21(1): 1-19.

［30］Zhu L, Yu X, Cao T, et al. Immune cell membrane-based biomimetic nanomedicine for treating cancer metastasis［J］. Acta Pharmaceutica Sinica B, 2023, 13(6): 2464-2482.

［31］Rezaie J, Nejati V, Mahmoodi M, et al. Mesenchymal stem cells derived extracellular vesicles: a promising

nanomedicine for drug delivery system［J］. Biochemical Pharmacology, 2022, 203: 115167.
［32］Wang J, Zhu F, Li K, et al. pH-responsive mesoporous Fe2O3-Au nanomedicine delivery system with magnetic targeting for cancer therapy［J］. Medicine in Novel Technology and Devices, 2022, 15: 100127.
［33］奉国和，刘任铧，邓伟伟. 高价值专利影响因素分析及主题挖掘——以纳米药物领域为例［J］. 图书馆论坛. 2024，44（7）：48-60.

Prediction and Verification of Emerging Topics from the Perspective of Dual Evolution of Indicators and Themes: The Case of Nanomedicine

Qiu Jing　Feng Guohe　Peng Kailin　Liu Renhua

(School of Economics & Management, South China Normal University, GuangZhou 510006, China)

Abstract:［**Purpose/Significance**］ Predicting emerging themes in a certain field is crucial for government departments to optimize resource allocation, enterprises to adjust their strategic direction, and researchers to track research trends and hotspots.［**Method/Process**］ The deep learning model BERTopic is used for theme extraction to predict emerging themes from two perspectives: one is to construct the novelty, growth and influence indicator system and improve the influence measurement indicators under the indicator evolution perspective, and the other is to carve the theme evolution relationship graph based on theme similarity under the theme evolution perspective. The accuracy and scientificity of emerging theme prediction are double-validated from the dimensions of theme evolution and data review.［**Result/Conclusion**］ The dual evolution perspective based on multiple indicators and themes can predict the emerging themes more scientifically and effectively than a single perspective.

Keywords: Emerging themes prediction; BERTopic; Evolutionary trends; Emerging index; Nanomedicine

（本文责编：王秀玲）